Perceptual-cognitive expertise when refereeing the scrum in rugby union

Abstract

Compared to sports performers, relatively little is known about how sports officials make decisions at a perceptual-cognitive level. Thus, this study examined the decision-making accuracy and gaze behaviour of rugby union referees of varying skill levels while reviewing scrum scenarios. Elite (*n* = 9) and trainee (*n* = 9) referees, as well as experienced players (*n* = 9), made decisions while watching ten projected scrum clips and wearing a mobile eye-tracking system. Decision-making accuracy and gaze behaviour were recorded for each scrum. The elite and trainee referees made more accurate decisions than the players, and differences in gaze behavior were observed. The elite and trainee referees displayed lower search rates, spent more time fixating central-pack (i.e., front rows, binds, and contact point) and less time fixating outer-pack (e.g., second rows) and non-pack (e.g., other) locations, and exhibited lower entropy than the players. While search rate failed to predict decision-making accuracy, the time spent fixating central-, outer-, and non-pack locations, as well as entropy, were significant predictors. The findings have implications for training perceptual-cognitive skill among sports officials.

Introduction

Effective decision-making requires the integration of existing knowledge with unfolding information in the performance environment. This ability to recognise and process the most relevant information, at the right time, in order to select an appropriate response is known as perceptual-cognitive expertise (Marteniuk, 1976). While abundant research has shown how sports performers make decisions at a perceptual-cognitive level (Mann, Williams, Ward & Janelle, 2007), relatively few studies have focused on sports officials or referees. This is surprising given that perception is fundamental to officiating (MacMahon, Mascarenhas, Plessner, Pizzera, Oudejans, & Raab, 2014), and referees need to make multiple decisions per minute (e.g., three to four per minute in soccer; Helsen & Bultynck, 2004) that can influence match outcomes, enforce the laws of the game, maintain ‘fair play’, and protect sports performers from injury. Indeed, perceptual-cognitive expertise is arguably more important for sports officials than performers, with referees required to perceive fast-paced actions from multiple performers in a limited time frame, categorise these actions as legal or illegal based on information retrieved from long-term memory, store this information in working memory, and use this information to make decisions which are heavily scrutinized by performers, coaches, and spectators (Plessner & Haar, 2006). Thus, this study used eye-tracking technology to offer a better understanding of the gaze behaviours used by referees of varying skill levels when making decisions under time pressure. By illuminating underlying attentional processes, it is hoped that the findings might help guide the education of the next generation of referees.

Knowing where and when to look, and being able to identify and process task-relevant information while ignoring less relevant information, is crucial for optimal decision-making (Williams, Davids, & Williams, 1999). Research using an expert-novice paradigm has supported this assertion, demonstrating that in comparison to novice sports performers, experts tend to employ gaze behaviour characterised by fewer fixations of a longer duration towards key perceptual cues (Mann et al., 2007). This finding is consistent with the information-reduction hypothesis (Haider & Frensch, 1999), which suggests that through amassed experience, experts allocate attention selectively towards task-relevant areas of the display and neglect task-redundant areas. However, it is important to note that perceptual-cognitive processes are considered highly task-dependent (Williams, Davids, Burwitz, & Williams, 1993). For example, in contrast to the typical result noted above, research using team-based decision-making tasks (e.g., 11 vs. 11 defensive soccer situations) has revealed that experts display more fixations of a shorter duration towards task-relevant areas (e.g., Roca, Ford, McRobert, & Williams, 2011; Vaeyens, Lenoir, Williams, & Philippaerts, 2007; Ward, Williams, & Bennett, 2002). Indeed, a meta-analysis by Gegenfurtner et al. (2011) highlighted that gaze behaviors can differ based on task characteristics such as dynamics, with experts more likely to employ a strategy consisting of more fixations of a shorter duration during relatively dynamic tasks, but fewer fixations of a longer duration during comparatively static tasks (Gegenfurtner, Lehtinen, & Saljo, 2011).

To date, relatively few studies have extended the expertise paradigm to the gaze behaviour of sports officials (MacMahon et al., 2014). Specifically, two studies have been conducted with reactors, or officials that monitor a low to medium number of cues and have little interaction with sports performers (MacMahon & Plessner, 2008). Both studies found that while reviewing offside situations, higher- and lower-level assistant soccer referees did not differ in terms of the number and duration of fixations, or the time spent fixating the passer and offside line (Catteuw, Helsen, Gilis, Van Roie, & Wagemans, 2009; Schnyder, Koedijker, Kredel, & Hossner, 2017). In addition, two studies have been performed with monitors, or officials that assess a high number of cues but have little interaction with sports performers (MacMahon & Plessner, 2008). Both studies reported that, while assessing single-gymnast routines, higher- and lower-level judges did not differ in terms of the number and duration of fixations, but that higher-level judges fixated more on the upper body (Bard, Fleury, Carriere, &

Halle, 1980; Pizzera, Moller, & Plessner, 2018). Finally, two studies have been conducted with interactors, or officials who monitor a high number of cues and regularly interact with sports performers (MacMahon & Plessner, 2008). Using ice hockey and soccer referees, these studies found that higher- and lower-level referees did not differ in the number and duration of fixations when making decisions (Hancock & Ste-Marie, 2013; Spitz, Put, Wagemans, Williams, & Helsen, 2016). However, compared to the lower-level soccer referees, the higher-level soccer referees spent more time fixating the body part of the attacking player involved in the infringement during open play situations (Spitz et al., 2016).

Taken together, this research suggests that higher-level referees may use similar visual search behaviours (i.e., number and duration of fixations) to lower-level referees, but fixate more relevant and information-rich locations, consistent with the information-reduction hypothesis (Haider & Frensch, 1999). Indeed, this effect seems particularly prominent among monitors and interactors who are often required to monitor more perceptual cues than reactors. However, further research is needed to substantiate this notion, given the limited number of studies conducted to date, and the limitations inherent within these studies. First, prior research has typically used tasks that involve less complex ‘matter of fact’ decisions with relatively few sports performers, and thus perceptual cues and possible infractions (e.g., offside or not; Catteuw et al., 2009). Given that expertise differences in perceptual-cognitive skill seem more likely to emerge in tasks of higher complexity (Gegenfurtner et al., 2011), future research should adopt tasks that require more difficult and ambiguous decisions with multiple performers (e.g., ‘matter of opinion’; MacMahon & Plessner, 2008). Second, existing work has tended to employ relatively dynamic decision-making tasks (e.g., handsprings forward with a half turn on/half turn off the vault; Pizzera et al., 2018). Given that gaze behaviors are sensitive to task constraints such as dynamics (Gegenfurtner et al., 2011), future research should also use relatively static (or less dynamic) tasks to aid our understanding of the perceptual processes of sports officiating.

Third, previous research has typically used decision-making tasks in which referees watch brief video clips (e.g., 4 s in duration; Hancock & Ste-Marie, 2013), on relatively small screens (e.g., 17-inch; Spitz et al., 2016). Research has shown that expertise differences in gaze behaviour are more likely to emerge during more realistic decision-making tasks that more closely resemble the natural performance environment (Kredel, Vater, Klostermann, & Hossner, 2017). Thus, although it is difficult to measure gaze behaviour *in situ*, future research should use decision-making tasks that include longer video clips projected onto a larger screen (Al-Abood, Bennett, Hernandez, Ashford, & Davis, 2002; MacMahon & Plessner, 2008). Fourth, previous research has tended to examine only two experimental groups of referees (e.g., national- and local-level; Bard et al., 1980). To understand more about how task-specific perceptual-cognitive expertise may develop, it would be interesting for future research to include a third ‘novice’ group consisting of sports performers who know the laws of the game, but have no prior refereeing experience. Fifth, while research has started to consider the importance of fixation location, or what information referees are utilising to make decisions (e.g., Spitz et al., 2016), future research should incorporate more sophisticated analyses of the top-down nature of gaze behavior (e.g., entropy; Shannon, 1948), and use statistical techniques to determine which gaze variables are most important for proficient decision-making (e.g., regression analyses).

The present study aimed to address these limitations and examine the decision-making accuracy and gaze behaviours of elite and trainee rugby union referees, as well as players, while reviewing scrum scenarios. The rugby union scrum offered a relatively more complex and less dynamic (or more static) decision-making task when compared to the tasks employed previously, with referees required to monitor the actions of multiple sports performers as they unfold, and select a fairly ambiguous or ‘matter of opinion’ decision (i.e., play on, reset, or penalty against attack/defense). Based on existing research (e.g., Spitz et al., 2016), it was predicted that the elite group would make more accurate decisions, fixate more on particular areas of the display (e.g., central- rather than outer- or non-pack locations), and display lower entropy (i.e., gaze distributed or spread less across locations) than the trainee and player groups, and that a similar differentiation would occur when comparing the trainee and player groups. It was also predicted that no significant differences would exist between the elite and trainee groups in terms of the number and duration of fixations (i.e., search rate); however, consistent with the findings of previous research using team-based decision-making tasks most comparable to the task employed in this study (e.g., Roca et al., 2011), the elite and trainee groups were expected to display more fixations of a shorter duration (i.e., higher search rate) than the player group. Finally, given these hypotheses, percentage viewing time to key locations and entropy were expected to predict decision-making accuracy, while search rate was not.

Method

Participants

Twenty-seven rugby union referees and players from the United Kingdom were recruited based on their previous experience and competitive level. The first group consisted of elite referees (n = 9; M_{age} = 30 years, SD = 6), who were refereeing the highest division of professional rugby (i.e., Premiership), many of whom were refereeing, or had refereed, at international level. The second group comprised trainee referees (n = 9; M_{age} = 20 years, SD = 1), who were from a University-based academy who refereed at lower competitive levels (i.e., county), but had little experience refereeing professionally. The third group consisted of players (n = 9; M_{age} = 33 years, SD = 5), who had never refereed, but were experienced in playing competitive rugby ($M_{experience}$ = 16 years, SD = 8). This study received institutional ethical approval and all participants provided informed consent.

Equipment and task

Gaze behaviour was measured using a SensoMotoric Instruments (SMI; Boston, MA) mobile eye-tracker. This lightweight (76 g) binocular system uses dark pupil tracking to calculate point of gaze and record the visual scene at a temporal resolution of 30 Hz and a spatial resolution of 0.5°. Gaze was viewed in real time by the researcher using a laptop (Lenovo, ThinkPad) installed with iViewETG software. Participants were connected to the laptop via a 1.80 m usb cable, and the researcher and laptop were located behind the participant to minimise distractions. The gaze data was recorded for subsequent offline analysis. The task required participants to make decisions regarding possible infractions while watching video clips of different scrum scenarios projected onto a 2.10 m or 83-inch (diagonally measured) screen using a LCD projector (Hitachi, CP-X4015WN 3LCD). Participants stood approximately 2.50 m from the screen, subtending a 45° visual angle. After each video clip, the screen went black for 10 s while participants verbalised their decision.

Video clips

Several steps were undertaken to design the video clips (as Hancock & Ste-Marie, 2013). First, a referee manager from the Rugby Football Union Professional Games Match Officials' Team (RFU PGMOT) provided video footage of scrum scenarios from televised rugby matches from their archive. These matches were from the highest professional leagues and competitions in club and international rugby. Second, this footage was edited and assembled using iMovie software (Apple Inc., United States), producing video clips from more of an 'in-game' (or assistant referees') perspective to enhance the representativeness of the task, and thus the likelihood of revealing expertise differences in gaze behaviour (Dicks, Davids, & Button, 2009). Each video clip started before the "set" call, and while the match referee was present in each video clip, each video clip was edited to ensure it finished before the referee revealed their decision, thus preventing the in-game referee from impacting decision-making. Third, the lead researcher and referee manager reviewed the 20 edited video clips and selected the final 10 video clips using criteria including video length and clarity, quality of vantage point or line of sight, type of infraction, and decision ambiguity. This resulted in 10 video clips of scrum scenarios ranging from 5 to 25 s in duration (*M* = 11.00 s, *SD* = 5.73). Each video clip contained only one possible infraction. The video clips were played with no sound to remove the influence of crowd, commentator, and player noise (Nevill, Balmer, & Williams, 2002).

Procedure

Participants first read an information sheet before providing written informed consent. Next, participants were fitted with the mobile eye-tracker, which was calibrated using a 9-point grid. Participants were then provided with a standardised and detailed verbal explanation of the task, before watching one video clip as a familiarisation. Participants were instructed to watch each video clip before verbalising their decision as quickly as possible once the screen went black. For each scrum scenario, participants made one of four decisions: (1) play on (i.e., no penalty), (2) reset, (3) penalty against attacking team (i.e., team putting into the scrum), or (4) penalty against defending team (i.e., team not putting into the scrum). To ensure all participants understood the task, after the familiarisation video clip, participants were asked if they had any questions. Subsequently, participants watched the 10 scrum video clips and stated their decisions while gaze behaviour and decisions were recorded. None of the video clips were replayed, and no feedback was given to participants between the clips. Finally, the mobile eye-tracker was removed, and participants were debriefed and thanked for their participation.

Measures

Decision-making accuracy

Two referee managers from the RFU PGMOT watched each scrum scenario before coming to an agreement on the correct (or reference) decision. ‘Play on’ and ‘reset’ were deemed the correct decision for two video clips each, and ‘penalty against attacking team’ and 'penalty against the defending team' were agreed as the reference decision for three video clips each. Decision-making accuracy was calculated as the total number of decisions (displayed as a percentage) that were in correspondence with the reference decision (as Spitz et al., 2016).

Gaze behaviour

A video recording containing each participant's eye movements (via a gaze cursor with a radius of 0.5°) was downloaded using BeGaze software (www.smivision.com). These videos were then analysed frame-by-frame across the entirety of each scrum scenario using Quiet Eye Solutions software (www.quieteyesolutions.com). A fixation was defined as a gaze that was maintained on a location within 1° of visual angle for a minimum of 120 ms (Vickers, 2007). Three gaze measures were assessed for each of the 10 video clips, and averaged across scrum scenarios: (1) search rate, (2) percentage viewing time to key locations, and (3) entropy. Search rate was calculated by dividing the total number of fixations by the total duration of fixations towards all key locations (in seconds; as Nibbeling, Oudejans, & Daanen, 2012). Percentage viewing time referred to the percentage of total viewing time spent fixating each location (as Roca et al., 2011). Following discussions with the referee manager, 14 possible fixation locations were identified including (1) attacking front row, (2) attacking second row, (3) attacking back row, (4) attacking scrum half, (5) defensive front row, (6) defensive second row, (7) defensive back row, (8) defensive scrum half, (9) contact point (i.e., point where the front rows met or contacted one another), (10) binds (i.e., where the front rows held or bound onto one another), (11) tunnel (i.e., gap between the lower bodies of the front rows in which the ball would be fed), (12) ball, (13) referee, and (14) other (e.g., crowd).

To simplify analyses, the fixation locations noted above were combined. Specifically, following discussions with the referee manager, three fixation locations were created: (1) central-pack, (2) outer-pack, and (3) non-pack. Central-pack comprised attacking front row, defensive front row, contact point, and binds. Outer-pack consisted of attacking second row, attacking back row, attacking scrum half, defensive second row, defensive back row, and defensive scrum half. Non-pack comprised tunnel, ball, referee, and other. These locations are displayed in Figure 1. Finally, entropy was calculated. Entropy refers to the uncertainty within a system, indicating the variability of gaze behaviour. While different measures of entropy exist (e.g., Allsop & Gray, 2014), Shannon entropy derives from information theory (Shannon, 1948), and expresses the information contained within a probability distribution in 'bits'. It is calculated from the state space of the system (all possible outcomes) and the relative probabilities of all elements in that state-space. Elements were defined as 13 key locations around the scrum (e.g., contact point, binds) plus 'other' (e.g., crowd). Entropy was calculated as the sum of the logarithm of all probabilities in the given state space, $\mathrm{H}(x) = -\sum_{i=1}^{n} P(x_i) log_b P(x_i)$, (Shannon, 1948). In short, the probability of fixating each location was calculated for each group, before applying the above formula to those probabilities. In the present study, lower entropy values therefore reflected gaze behaviour that was focused on particular fixation locations, rather than distributed or spread evenly across all locations.

**** Figure 1 near here ****

Statistical analyses

A series of one-way ANOVAs with post hoc LSD *t*-tests were used to examine between-group differences in experience, decision-making accuracy, search rate, and entropy. Percentage viewing time for the fixation locations was analysed using a two-way ANOVA with group (elite vs. trainee vs. player) as the between-subjects factor and fixation location (central-vs. outer- vs. non-pack) as the within-subjects factor. Significant main and interaction effects were followed up with post-hoc one-way ANOVAs and LSD *t*-tests. In all ANOVAs in which the sphericity assumption was violated, the degrees of freedom were corrected using the Greenhouse-Geisser correction procedure. Effect sizes were calculated as partial eta-squared (η_p^2), with values of 0.01, 0.06, and 0.14 interpreted as small, medium, and large effects, respectively. Finally, a series of bivariate regression analyses were conducted to examine if search rate, percentage viewing time to central-, outer-, and non-pack locations, or entropy, predicted a significant amount of variance in decision-making accuracy. These regression analyses were based on mean values for each participant's performance and gaze metrics. A *p*-value of less than .05 was considered statistically significant (Field, 2013), and all analyses were conducted using IBM SPSS statistical program version 22.

Results

Experience

There was a significant difference between the groups, $F(2, 24) = 21.69$, $p < .001$, $\eta_p^2 = .64$. The elite group reported greater refereeing experience than both the trainee ($p < .001$) and player ($p < .001$) groups. Furthermore, the trainee group reported more refereeing experience than the player group ($p = .041$). The referee experience data are presented in Table 1.

Decision-making accuracy

There was a significant difference between the groups, $F(2, 24) = 5.21$, $p = .013$, $\eta_p^2 = .30$. The player group made significantly fewer correct decisions than the elite ($p = .027$) and trainee ($p = .005$) groups. There was no significant difference between the elite and trainee groups ($p = .475$). The decision-making accuracy data are presented in Table 1.

Search rate

There was a significant difference between the groups, $F(2, 24) = 9.07$, $p = .001$, $\eta_p^2 = .43$. The player group exhibited a significantly higher search rate than the elite group ($p < .001$), and a marginally higher search rate than the trainee group ($p = .063$). In addition, the trainee group displayed a significantly higher search rate than the elite group ($p = .030$). The search rate data are presented in Table 1.

Percentage viewing time

There were significant main effects for group, $F(2, 24) = 9.18$, $p = .001$, $\eta_p^2 = .43$, and fixation location, $F(1.33, 31.79) = 864.64$, $p < .001$, $\eta_p^2 = .97$, and a significant interaction effect, $F(2.65, 31.79) = 9.29$, $p < .001$, $\eta_p^2 = .44$. Follow-up between-subjects analyses revealed that the player group spent significantly less time fixating central-pack locations than the elite and trainee groups (*ps* = .002), with no significant difference between the elite and trainee groups ($p = .960$). Moreover, the player group spent significantly more time fixating outer-pack locations than the elite and trainee groups (*ps* = .001), with no significant difference between the elite and trainee groups ($p = .916$). Furthermore, the player group spent significantly more time fixating non-pack locations than the elite and trainee groups (*ps* ≤ .002), with no significant difference between the elite and trainee groups ($p = .174$). Follow-up within-subjects analyses revealed that all three groups spent significantly more time fixating central-pack locations than outer- and non-pack locations (all *ps* ≤ .004). The percentage viewing time data is presented in Table 1.

Entropy

There was a significant difference between the groups, $F(2, 24) = 5.23$, $p = .013$, $\eta_p^2 = .30$. The player group displayed significantly greater entropy than the elite ($p = .007$) and trainee ($p = .016$) groups. There was no significant difference between the elite and trainee groups ($p = .699$). The entropy data are presented in Table 1.

**** Table 1 near here ****

Regression analyses

Search rate did not account for a significant proportion of variance in decision-making accuracy, R^2 = .07, β = -.32, p = .103, *95% CI* = -30.09 to 2.95. However, percentage viewing time to central-pack, R^2 = .19, β = .47, p = .013, *95% CI* = 0.17 to 1.31, outer-pack, R^2 = .24, β = -.52, p = .006, *95% CI* = -1.81 to -0.34, and non-pack, R^2 = .14, β = -.42, p = .029, *95% CI* = -2.17 to -0.13, locations as well as entropy, R^2 = .13, β = -.41, p = .036, *95% CI* = -29.57 to -1.11, accounted for a significant proportion of variance in decision-making accuracy. These results suggest that more time fixating central-pack locations, less time fixating outer- and non-pack locations, and lower entropy, were associated with more accurate decisions.

**** Figures 2 and 3 near here ****

Discussion

Abundant research has highlighted how sports performers make decisions at a perceptual-cognitive level (Mann et al., 2007), however, comparatively little work has focused on sports officials (MacMahon et al., 2014). Thus, this study used eye-tracking technology to better understand the gaze behaviours used by referees of varying skill levels when making decisions under time pressure. As hypothesised, the groups differed in terms of decision-making accuracy, an effect that was largely driven by the player group making poorer decisions than the elite and trainee groups. Despite their playing involvement, the player groups' lack of refereeing experience might have meant that they did not possess the specific knowledge required to make effective decisions (Ericsson & Kintsch, 1995). Although it should be noted that the player group achieved a level of decision-making accuracy greater than would be expected by chance (i.e., 39%), suggesting an adequate understanding of the task, possibly owing to their previous experience playing in, and spectating, rugby union matches (Pizzera & Raab, 2012).

However, contrary to previous research revealing expertise differences in decision-making accuracy (e.g., Hancock & Ste-Marie, 2013; Spitz et al., 2016), the elite and trainee referees made decisions of similar accuracy. One possible explanation for this result might be that the trainee group had acquired enough refereeing experience (4 years on average) to develop the knowledge required to make appropriate decisions during scrum scenarios, with key factors other than decision-making distinguishing them from their elite counterparts (e.g., perceptual-cognitive skill in other scenarios such as rucks, or communication and player management; Cunningham, Simmons, Mascarenhas, & Redhead, 2014). Indeed, it is worth noting that the trainee referees were part of a University-based academy which has previously produced two elite referees, and thus received regular training and support on several aspects of refereeing including managing the scrum. Although this result was unexpected, it should be noted that not all research has revealed expertise differences (e.g., Bard et al., 1980), particularly when investigating officiating in rugby union (MacMahon & Ste-Marie, 2002). For example, Mascarenhas et al. (2005) found that rugby union officials ranked in the top-20 were as accurate as lower ranked (41^{st}-65^{th}) referees when making decisions during tackle scenarios (54% vs. 52%; Mascarenhas, Collins, & Mortimer, 2005). Another potential explanation could be that the elite referees underperformed because the frequency with which the decisions (i.e., play on, reset, and penalty) were presented during the task, differed to actual game demands. For instance, the correct decision was to award a penalty in 60% of the video clips, when a penalty is typically blown in ~42% of scrums (Six Nations Statistical Report, 2015). Indeed, there is growing awareness that expert referees make better use of such contextual information to support their decision-making relative to novice referees (e.g., previous decisions; Unkelbach & Memmert, 2008). However, it should be noted that little is currently known about how frequently penalties are awarded in the scrum at lower competitive levels, and thus more research is required before this explanation can be accepted or refuted.

While somewhat limited, existing research has found that sports officials of varying skill levels do not differ in terms of visual search behaviour, implying that higher-level referees might interpret or categorise visual information better to make more accurate decisions (e.g., Hancock & Ste-Marie, 2013; Spitz et al., 2016). Contrary to previous research, the elite group displayed a lower search rate characterised by fewer fixations of a longer duration than the trainee and player groups. This unexpected result might be attributable to the longer video clips and larger screen employed, making the decision-making task more akin to the natural environment (Al-Abood et al., 2002; MacMahon & Plessner, 2008). Indeed, it has been argued that expertise differences in gaze behaviour are more likely to emerge during more realistic tasks (Dicks et al., 2009; Kredel et al., 2017). Alternatively, this finding might be due to the scrum scenarios being relatively more static (or less dynamic) than the tasks employed previously (e.g., ice hockey open-play; Hancock & Ste-Marie, 2013). Indeed, compared to novices, expert sports performers generally display lower search rates when performing relatively static sporting tasks (e.g., tennis serve return; Gegenfurtner et al., 2011; Mann et al., 2007). Given that expertise differences in perceptual-cognitive skill seem more likely to emerge in more complex tasks (Gegenfurtner et al., 2011), a final explanation could be that the scrum scenarios were relatively more difficult and ambiguous than the tasks adopted previously, which have tended to involve ‘matter of fact’ decisions with few sports performers (e.g., offside or not; Catteuw et al., 2009). From the perspective of the information-reduction hypothesis (Haider & Frensch, 1999), this result might suggest that through experience, the elite referees have learnt to optimise the amount of information they process, neglecting task-redundant cues and selectively focusing on task-relevant information. Interestingly, search rate did not significantly predict decision-making accuracy and only accounted for 7% of variance (equating to a small to medium effect size; Cohen, 1992), implying that visual search might not determine decision-making proficiency.

Regardless of group, participants fixated more (70-85%) on central- (i.e., front rows, contact point, and binds), rather than outer- (i.e., second rows, back rows, and scrum halves) or non- (i.e., tunnel, ball, referee, and other) pack locations. While speculative, this gaze strategy, combined with the longer fixations, might indicate the use of a visual pivot, where foveal attention is focused centrally and peripheral vision is used to detect exterior cues and guide future eye movements (Williams & Elliott, 1999). Indeed, such a strategy might be beneficial given that information can be more readily extracted from peripheral vision when the eyes are stationary rather than moving (Motter & Simoni, 2008). Alternatively, rather than improving the referees ability to use peripheral vision to locate possible infractions in non-central locations (e.g., ball feed), this gaze strategy might simply reflect that central-pack locations contain the most important visual information needed to make decisions during scrums, with a higher proportion of infractions stemming from these areas (e.g., angle of front row; Six Nations Statistical Report, 2017). However, the eye-tracker employed in this study was unable to account for potential information pick-up from peripheral vision, and so future research should use more suitable technology to better elucidate the role of central and peripheral vision in sports officiating (e.g., gaze-contingent displays). Indeed, research among sport performers has shown that more highly skilled performers tend to make better use of both central and peripheral vision when making decisions (e.g., Ryu, Abernethy, Mann, Poolton, & Gorman, 2013).

Consistent with previous research (Spitz et al., 2016), the groups differed in terms of the time spent fixating different locations. Indeed, compared to the elite and trainee groups, the player group spent less time fixating central-pack locations, and more time fixating outer- and non-pack locations. Thus, as a result of their limited refereeing experience, the player group spent longer fixating outer- and non-pack locations, which might have prevented them from developing a complete mental ‘picture’ of the situation and focusing on more relevant information, potentially resulting in erroneous decisions (Ericsson & Kintsch, 1995; Haider & Frensch, 1999). However, in contrast to prior work (Spitz et al., 2016), no differences were observed between the elite and trainee groups. Although unexpected, this finding mirrored the decision-making accuracy results, implying that due to their previous experience officiating the scrum, both the elite and trainee referees were able to identify and focus on information from central regions of the display (e.g., front row binds), while ignoring information from outer- and non-pack locations (e.g., scrum halves), possibly leading to more accurate decisions, given that most scrum infractions emanate from these central locations (e.g., collapsing and binding, angle and wheeling, standing up; Six Nations Statistical Report, 2017). This was supported by the regression results, which showed that spending more time fixating central-pack locations, and less time fixating outer- and non-pack locations, was associated with more accurate decisions. Indeed, these variables accounted for between 14% and 24% of variance, and equated to medium to large effect sizes (Cohen, 1992).

This study was the first to investigate expertise differences in entropy within the domain of sports officiating or refereeing, revealing differences between the groups. As predicted, the player group displayed greater entropy compared to the elite and trainee groups. In other words, fixations were more evenly distributed across all locations for the player group relative to the elite and trainee groups (Ellis & Stark, 1986). Thus, possibly owing to their lack of referee-specific experience (Haider & Frensch, 1999), the player group distributed or spread their visual attention more widely across the different areas of the display (indicative of stimulus-driven or bottom-up control; Malcolm & Henderson, 2010), potentially compromising their decision-making, particularly if this resulted in key perceptual information being missed. Contrary to predictions, but in line with the decision-making accuracy results, the elite and trainee groups displayed similar entropy. This might suggest that the experience officiating the scrum that these referees have accrued enabled them to develop a more systematic visual search (Haider & Frensch, 1999), allowing them to focus on more critical aspects of the display while ignoring less relevant aspects (reflective of goal-directed or top-down attentional control; Malcolm & Henderson, 2010), potentially leading to superior decision-making. Consistent with this notion, lower entropy predicted more accurate decision-making, accounting for 13% of variance, and equating to a medium to large effect size (Cohen, 1992). However, this result should be interpreted cautiously given the trend (albeit non-significant) for the trainee group to display higher entropy, but make more accurate decisions, than the elite group.

Video-based training has been shown to benefit sports officials' decision-making (e.g., Mascarenhas, Collins, Mortimer, & Morris, 2005; Schweizer, Plessner, Kahlert, & Brand, 2011). Given the results of the regression analyses, individualised eye movement training programmes could be developed to help referees who struggle to officiate the scrum employ the visual search strategies that are associated with more accurate decision-making, including longer fixations towards central-pack locations (i.e., front rows, binds, contact point). Indeed, despite mixed evidence in sporting tasks (e.g., Abernethy, Schorer, Jackson, & Hagemann, 2012; Ryu, Kim, Abernethy, & Mann, 2014), such eye movement training has proved beneficial in non-sporting tasks including medical screening and fingerprint matching (Litchfield, Ball, Donovan, Manning, & Crawford, 2010; Roads, Mozer, & Busey, 2016). Despite this implication, several limitations and directions for future research should be noted. First, given that the video clips were edited from televised matches, it is possible that the referees had seen some of the scenarios before or even officiated in the matches. However, a referee will often officiate 25 matches per season with an average of 18 scrums per match, totaling 450 scrums. Thus, it is unlikely that the referees remembered each scrum, although future research should control for this by generating unique first-person footage (as Spitz et al., 2016). Indeed, such footage would also help improve the representativeness of the task employed in this study, which predominately adopted the perspective of the assistant referee. This is likely to be important given that different viewing perspectives have been shown to influence visual search behaviour (e.g., Mann, Farrow, Shuttleworth, & Hopgood, 2009).

Second, although the number of video clips employed in this study was consistent with previous research revealing differences in perceptual-cognitive expertise between professional and amateur soccer referees (i.e., Spitz et al., 2016), the relatively low number of trials utilised compared to previous research could be considered a limitation (e.g., Schnyder et al., 2017), preventing the emergence of differences in decision-making accuracy and visual search behaviour between the expert and trainee referees. Thus, researchers are encouraged to employ more trials in future decision-making tasks among sports officials (Kredel et al., 2017). Third, although the present study tried to employ a more realistic referee-specific decision-making task than previous research (e.g., more complex and dynamic clips; longer clips projected onto a larger screen), the task was still conducted in a laboratory rather than a naturalistic setting, limiting the representativeness of the task (e.g., shorter ‘lead in’, smaller visual angle). Given that gaze behaviours can differ between these contexts due to different task constraints (Dicks, Button, & Davids, 2010), future research should investigate the gaze behaviors employed by sports officials *in situ*, to better elucidate the perceptual-cognitive processes underlying expert decision-making (Dicks et al., 2009). Finally, although it offered an expedient marker of the distribution or spread of gaze, the measure of entropy used in this study did not take the time sequence of fixation locations or the dynamics of the scene into account. Thus, future work is encouraged to calculate other indices of entropy that better capture the timing and duration of critical events as well as time-related gaze behaviour (e.g., Allsop & Gray, 2014).

To conclude, this study examined the decision-making accuracy and gaze behaviours of rugby union referees of varying skill levels while assessing scrum scenarios. Compared to the players, the elite and trainee referees made more accurate decisions, displayed lower search rates, spent more time fixating central-pack and less time fixating outer- and non-pack locations, and exhibited less entropy. The findings highlight the gaze strategies that are associated with more accurate decision-making in scrum scenarios, and could therefore be incorporated into individualised training programmes aimed at improving the decision-making of referees.

Disclosure of Interest

The authors report no conflicts of interest.

References

Abernethy, B., Schorer, J., Jackson, R.C., & Hagemann, N. (2012). Perceptual training methods compared: The relative efficacy of different approaches to enhancing sport-specific anticipation. *Journal of Experimental Psychology: Applied*, *18*, 143-153. doi:10.1037/a0028452.

Al-Abood, S.A., Bennett, S.J., Hernandez, F.M., Ashford, D., & Davids, K. (2002). Effect of verbal instructions and image size on visual search strategies in basketball free throw shooting. *Journal of Sports Sciences*, *20*, 271-278. doi:10.1080/026404102317284817.

Allsop, J., & Gray, R. (2014). Flying under pressure: Effects of anxiety on attention and gaze behavior in aviation. *Journal of Applied Research in Memory and Cognition*, *3*, 63-71. doi:10.1016/j.jarmac.2014.04.010.

Bard, C., Fleury, M., Carriere, L., & Halle, M. (1980). Analysis of gymnastics judges' visual search. *Research Quarterly for Exercise and Sport*, *51*, 267-273. doi:10.1080/02701367.1980.10605195.

Catteeuw, P., Helsen, W., Gilis, B., Van Roie, E., & Wagemans, J. (2009). Visual scan patterns and decision-making skills of expert assistant referees in offside situations. *Journal of Sport and Exercise Psychology*, *31*, 786-797.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*, 155-159. doi:10.1037/0033-2909.112.1.155.

Cunningham, I., Simmons, P., Mascarenhas, D., & Redhead, S. (2014). Skilled interaction: Concepts of communication and player management in the development of sport officials. *International Journal of Sport Communication*, *7*, 166-187. doi:10.1123/IJSC.2013-0098.

Dicks, M., Davids, K., & Button, C. (2009). Representative task designs for the study of perception and action in sport. *International Journal of Sport Psychology*, *40*, 506-524.

Dicks, M., Davids, K., & Button, C. (2010). Examination of gaze behaviors under in situ and video simulation task constraints reveals differences in information pick-up for perception and action. *Attention, Perception, and Psychophysics*, *72*, 706-720. doi:10.3758/APP.72.3.706.

Ellis, S.R., & Stark, L. (1986). Statistical dependency in visual scanning. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *28*, 421-438.

Ericsson, K.A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211-245. doi:10.1037/0033-295X.102.2.211.

Field, A. (2013). *Discovering statistics using SPSS* (4th Ed). Washington, DC: Sage.

Gegenfurtner, A., Lehtinen, E., & Saljo, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, *23*, 523-552. doi:10.1007/s10648-011-9174-7.

Haider, H., & Frensch, P.A. (1999). Eye movement during skill acquisition: More evidence for the information reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *25*, 172-190. doi:10.1037/0278-7393.25.1.172.

Hancock, D.J., & Ste-Marie, D.M. (2013). Gaze behaviors and decision-making accuracy of higher- and lower-level ice hockey referees. *Psychology of Sport and Exercise*, *14*, 66-71. doi:10.1016/j.psychsport.2012.08.002.

Helsen, W.F., & Bultynck, J. (2004). Physical and perceptual-cognitive demands of top-class refereeing in association football. *Journal of Sports Sciences*, *22*, 179-189. doi:10.1080/02640410310001641502.

Kredel, R., Vater, C., Klostermann, A., & Hossner, E.J. (2017). Eye-tracking technology and the dynamics of natural gaze behavior in sports: A systematic review of 40 years of research. Frontiers in Psychology: Performance Science. doi:10.3389/fpsyg.2017.01845.

Litchfield, D., Ball, L., Donovan, T., Manning, D.J., & Crawford, T. (2010). Viewing another person's eye movements improves identification of pulmonary nodes in chest x-ray inspection. *Journal of Experimental Psychology: Applied*, *16*, 251-262. doi:10.1037/a0020082.

MacMahon, C., & Plessner, H. (2008). The sport official in research and practice. In D. Farrow, J. Baker, & C. MacMahon (Eds.). *Developing sport expertise: Lessons from theory and practice* (pp. 172-190). New York, NY: Routledge.

MacMahon, C., & Ste-Marie, D.M. (2002). Decision-making by experienced rugby referees: Use of perceptual information and episodic memory. *Perceptual and Motor Skills*, *95*, 570-572. doi:10.2466/pms.2002.95.2.570.

Malcolm, G.L., & Henderson, J.M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision*, *10*, 1-11. doi:10.1167/10.2.4.

Mann, D.L., Farrow, D., Shuttleworth, R., & Hopwood, M. (2009). The influence of viewing perspective on decision-making and visual search behavior in an invasive sport. *International Journal of Sport Psychology*, *40*, 546-564.

Mann, D.T.Y., Williams, A.M., Ward, P., & Janelle, C.M. (2007). Perceptual-cognitive expertise in sport: A meta-analysis. *Journal of Sport and Exercise Psychology*, *29*, 457-478.

Marteniuk, R.G. (1976). *Information processing in motor skills*. New York: Holt, Rinehart, and Winston.

Mascarenhas, D.R.D., Collins, D., & Mortimer, P. (2005). The accuracy, agreement and coherence of decision-making in rugby union officials. *Journal of Sport Behavior*, *28*, 253-271.

Mascarenhas, D.R.D., Collins, D., Mortimer, P.W., & Morris, B. (2005). Training accurate and coherent decision-making in rugby union referees. *The Sport Psychologist*, *19*, 131-147. doi:10.1123/tsp.19.2.131.

Motter, B.C., & Simoni, D.C. (2008). Changes in the functional visual field during search with and without eye movements. *Vision Research*, *48*, 2382-2393. doi:10.1016/j.visres.2008.07.020.

Nevill, A., Balmer, N., & Williams, A.M. (2002). The influence of crowd noise and experience upon refereeing decisions in football. *Psychology of Sport and Exercise*, *3*, 261-272. doi:10.1016/s1469-0292(01)00033-4.

Pizzera, A., Moller, C., & Plessner, H. (2018). Gaze behavior of gymnastics judges: Where do experienced judges and gymnasts look while judging? *Research Quarterly for Exercise and Sport*, *89*, 112-119. doi:10.1080/02701367.2017.1412392.

Pizzera, A., & Raab, M. (2012). Perceptual judgements of sports officials are influenced by their motor and visual experience. *Journal of Applied Sport Psychology*, *24*, 59-72. doi:10.1080/10413200.2011.608412.

Plessner, H., & Haar, T. (2006). Sports performance judgments from a social cognition perspective. *Psychology of Sport and Exercise*, *7*, 555-575. doi:10.1016/j.psychsport.2006.03.007.

Nibbeling, N., Oudejans, R.R.D., & Daanen, H.A.M. (2012). Effects of anxiety, a cognitive secondary task, and expertise on gaze behavior and performance in a far aiming task. *Psychology of Sport and Exercise*, *13*, 427-435. doi:10.1016/j.psychsport.2012.02.002.

Roads, B., Mozer, M.C., & Busey, T.A. (2016). Using highlighting to train attentional expertise. *PLoS ONE*, *11*, 1-24. doi:10.1371/journal.pone.0146266.

Roca, A., Ford, P.R., McRobert, P.R., & Williams, A.M. (2011). Identifying the processes underpinning anticipation and decision-making in a dynamic time-constrained task. *Cognitive Processing*, *12*, 301-310. doi:10.1007/s10339-011-0392-1.

Ryu, D., Abernethy, B., Mann, D.L., Poolton, J.M., & Gorman, A.D. (2013). The role of central and peripheral vision in expert decision-making. *Perception*, *42*, 591-607. doi:10.1068/p7487.

Ryu, D., Kim, S., Abernethy, B., & Mann, D.L. (2014). Guiding attention aids the acquisition of anticipatory skill in novice soccer goalkeepers. *Research Quarterly for Exercise and Sport*, *84*, 252-262. doi:10.1080/02701367.2013.784843.

Schnyder, U., Koedijker, J.M., Kredel, R., & Hossner, E.J. (2017). Gaze behaviour in offside decision-making in football. *German Journal of Exercise and Sport Research*, *47*, 103-109. doi:10.1007/s12662-017-0449-0.

Schweizer, G., Plessner, H., Kahlert, D., & Brand, R. (2011). A video-based training method for improving soccer referees' intuitive decision-making skills. *Journal of Applied Sport Psychology*, *23*, 429-442. doi:10.1080/10413200.2011.555346.

Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379-423.

Six Nations Statistical Report (2015). Retrieved March 26th, 2018, from http://pulse-static-files.s3.amazonaws.com/test/worldrugby/document/2015/04/20/3ef09898-c8a1-4043-b060-75b043138015/150417_RJ_6_NATIONS_STATISTICAL_REPORT.pdf.

Six Nations Statistical Report (2017). Retrieved March 26th, 2018, from http://playerwelfare.worldrugby.org/content/getfile.php?h=7e525c749d663c9a901be0ba413b963e&p=pdfs/gameanalysis/70.pdf&d=6_Nations_analysis_2017.

Spitz, J., Put, K., Wagemans, J., Williams, A.M., & Helsen, W.F. (2016). Visual search behaviors of association football referees during assessment of foul play situations. *Cognitive Research: Principles and Implications*, *1*, 1-11. doi:10.1186/s41235-016-0013-8.

Unkelbach, C., & Memmert, D. (2008). Game-management, context-effects, and calibration: The case of yellow cards in soccer. *Journal of Sport and Exercise Psychology*, *30*, 95-109. doi:10.1123/jsep.30.1.95.

Vaeyens, R., Lenoir, M., Williams, A.M., & Philippaerts, R.M. (2007). Mechanisms underpinning successful decision-making in skilled youth soccer players: An analysis of visual search behaviors. *Journal of Motor Behavior*, *39*, 395-408. doi:10.3200/JMBR.39.5.395-408.

Vickers, J.N. (2007). *Perception, cognition, and decision training: The quiet eye in action*. Champaign, IL: Human Kinetics.

Ward, P., Williams, A.M., & Bennett, S.J. (2002). Visual search and biological motion perception in tennis. *Research Quarterly for Exercise and Sport*, *73*, 107-112. doi:10.1080/02701367.2002.10608997.

Williams, A.M., Davids, K., Burwitz, L., & Williams, J.G. (1993). Visual search and sports performance. *The Australian Journal of Science and Medicine in Sport*, *25*, 147-204.

Williams, A.M., Davids, K., & Williams, J.G. (1999). *Visual perception and action in sport*. London: E & FN Spon.

Williams, A.M., & Elliott, D. (1999). Anxiety, expertise, and visual search strategy in karate. *Journal of Sport and Exercise Psychology*, *21*, 362-375. doi:10.1123/jsep.21.4.362.

Table 1. Mean (standard deviation) experience, decision-making accuracy, search rate, entropy, and percentage viewing time data for scrum scenarios.

	Group		
	Elite	Trainee	Player
Experience (years)	11.67 (6.26)	3.89 (2.16)†	0.00 (0.00)*
Decision-making accuracy (%)	53.33 (14.14)	57.78 (10.93)	38.89 (13.64)*
Search rate (fixations per s)	1.64 (0.29)	1.94 (0.34)†	2.20 (0.19)*
Entropy (bits)	2.73 (0.37)	2.79 (0.39)	3.21 (0.25)*
Percentage viewing time			
Central-pack (%)	85.11 (5.21)	84.93 (7.22)	72.36 (9.76)*
Outer-pack (%)	12.84 (5.55)	12.56 (4.64)	22.67 (6.30)*
Non-pack (%)	2.77 (0.86)	5.42 (4.45)	11.85 (5.28)*

Note. * Significantly different from elite and trainee groups, † Significantly different from elite group

Figure 1. Example of a scrum scenario with a visualisation of the key fixation locations (white = central-pack, black = outer-pack, grey = non-pack).

Figure 2. Regression equation (with 95% CI) for decision-making accuracy (% correct) and (A) search rate (fixations per second) or (B) entropy (bits).

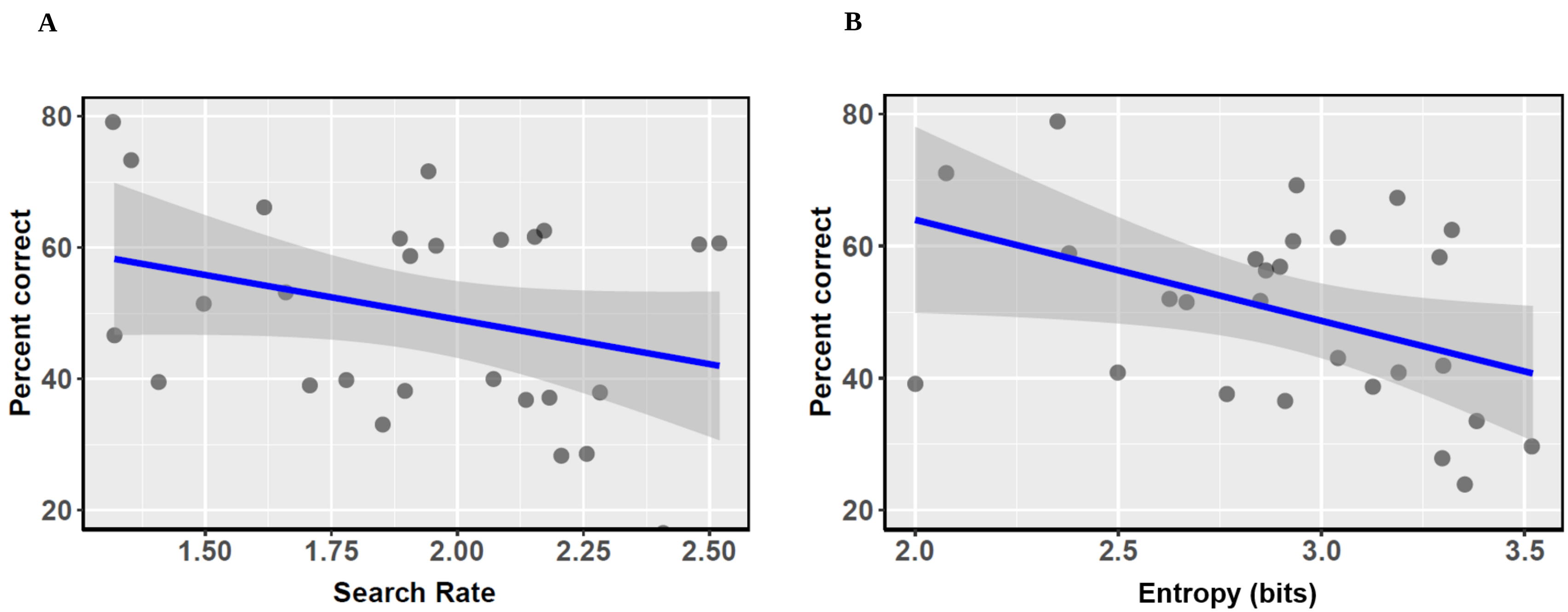

Figure 3. Regression equation (with 95% CI) for decision-making accuracy (% correct) and time spent viewing (A) central-, (B) outer-, or (C) non-pack locations (%).